APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/771,008	01/26/2001	Dallan W. Quass	20192-2.00	6345

			EXAMINER
20350	7590	12/04/2003	HOLMES, MICHAEL B

TOWNSEND AND TOWNSEND AND CREW, LLP
TWO EMBARCADERO CENTER
EIGHTH FLOOR
SAN FRANCISCO, CA 94111-3834

ART UNIT	PAPER NUMBER
2121	

DATE MAILED: 12/04/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/771,008	Applicant(s) QUASS ET AL.
	Examiner Michael B. Holmes	Art Unit 2121

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _(3)_ MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _26 January 2001_ .

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-15_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-15_ is/are rejected.

7)☒ Claim(s) _1-15_ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _26 January 2000_ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

    If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

        1.☐ Certified copies of the priority documents have been received.

        2.☐ Certified copies of the priority documents have been received in Application No. _____ .

        3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

    a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☒ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: .

UNITED STATES PATENT AND TRADEMARK OFFICE

P.O. Box 1450, Alexandria, Virginia 22313-1450 – WWW.USPTO.GOV

Examiner's Detailed Office Action

1.   This action is responsive to application **09/771,008**, filed **January 26, 2001**.

2.   **Claims 1-15** have been examined.

Information Disclosure Statement

3.   Applicant is respectfully remind of the Duty to disclose 37 C.F.R. 1.56 all pertinent

information and material pertaining to the patentability of applicant's claimed invention, by

continuing to submitting in a timely manner PTO-1449, Information Disclosure Statement (IDS)

with the filing of applicant's of application or thereafter.

Drawings

4.   The formal drawings have been reviewed by the United States Patent & Trademark

Office of Draftperson's Patent Drawings Review. Form PTO-948 has been provided.

Specification Objection

5.   Examiner objects to the specification as failing to adequately teach applicant's claimed invention. Specifically, the specification lacks the technical detailed that is normally associated with such an invention. Moreover, the specification juxtaposed the drawings do not adequately disclosed a detailed technical analysis of applicant's invention. On page 9, item 104, is lined out.

Claim Objection

6.   Examiner objects to the claim language, and has found claims 1-15 to be very, very confusing.

Claim Interpretation

7.   Office personnel are to give claims their **"broadest reasonable interpretation"** in light of the supporting disclosure. *In re Morris*, 127 F.3d 1048, 1054-55, 44 USPQ2d 1023, 1027-28 (Fed. Cir. 1997). Limitations appearing in the specification but not recited in the claim are not read into the claim. *In re Prater*, 415 F.2d 1393, 1404-05, 162 USPQ 541, 550-551(CCPA 1969). See *also *In re Zletz*, 893 F.2d 319, 321-22, 13 USPQ2d 1320, 1322(Fed. Cir. 1989) ("During patent examination the pending claims must be interpreted as broadly as their terms reasonably allow. . . . The reason is simply that during patent prosecution when claims can be amended, ambiguities should be recognized, scope and breadth of language explored, and clarification imposed. . . . An essential purpose of patent examination is to fashion claims that are precise, clear, correct, and unambiguous. Only in this way can uncertainties of claim scope be

removed, as much as possible, during the administrative process."). *see* MPEP § 2106

Claim Rejections - 35 USC § 103

8.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or
> described as set forth in section 102 of this title, if the differences between the subject
> matter sought to be patented and the prior art are such that the subject matter as a
> whole would have been obvious at the time the invention was made to a person having
> ordinary skill in the art to which said subject matter pertains. Patentability shall not be
> negatived by the manner in which the invention was made.

9.   **Claims 1-15** are rejected under 35 U.S.C. 103(a) as being unpatentable over

Chakrabarti et al. (USPN 6,389,436 B1), Filed Dec. 15, 1997; Date of Patent: May 14, 2002

in view of

Type Classification of Semi-Structured Documents, Markus Tresch; Neal Palmer; Allen

Luniewski; (hereinafter referred to as "*Tresch et al.*") Proceedings of the 21st VLDB

Conference, Zurich, Switzerland, (1995).

Regarding Claim 1:

Chakrabarti et al. teaches,

In a data processing system, a method for creating a database from information found on a

plurality of web pages using a first classifier and a second classifier, said method comprising:

a) defining first regularities and second regularities, said first regularities being patterns which

are expected to be found in information in said web pages, and said second regularities being

patterns which are not expected to be found in all said web pages **[(col. 6, line 33-42** "*The*

neighborhood could be of any size, for example, even including all of the documents available on

the Web. The neighborhood of documents, for a radius-one neighborhood, includes all documents to which hyperlinks (i.e., citations to and from other documents) in the new document point (i.e., out-links) and all documents containing hyperlinks that point to the new document (i.e., in-links); for a radius-two neighborhood, documents that in-link or out-link to documents that the new document in-links or out-links to are included; etc.")];

b) initially providing descriptions of said first regularities to a working database **[(col. 5, line 46-53** *"The present invention is usually (although not necessarily) implemented by a computer program known as a hypertext classifier 110 and an associated database 112 (both of which may be permanently stored on a data storage device coupled to the server 106). The hypertext classifier 110 classifies documents that contain hyperlinks, and more specifically, it classifies documents on the Internet 104 using a radius-one or radius-two neighborhood.")]*; thereafter to

c) training a first classifier in said working database using said first regularities **[("col. 6, line 33-40** *"The neighborhood could be of any size, for example, even including all of the documents available on the Web. The neighborhood of documents, for a radius-one neighborhood, includes all documents to which hyperlinks (i.e., citations to and from other documents) in the new document point (i.e., out-links) and all documents containing hyperlinks that point to the new document (i.e., in-links); ")]*;

d) identifying a candidate subset of the web pages expected to have said second regularities **[(col. 6, line 40-42** *"for a radius-two neighborhood, documents that in-link or out-link to documents that the new document in-links or out-links to are included; etc.")]*; thereafter

e) tentatively identifying and tagging, in said candidate subset of the web pages, elements having said first regularities, by using said first classifier to obtain first tentative labels **[(col. 6, line**

46-53 *"Next, the hypertext classifier 110 invokes a text-based classifier. The text-based classifier assigns a probability vector to each document. The probability vector contains a component for each possible class into which the document can be classified. For example, for five classes, a probability vector would have five components, each component indicating the probability that the document would fall under the associated class.")]*;

g) tentatively identifying elements having specific combinations of said first regularities and said second regularities using said first classifier and said second classifier to obtain second tentative labels for said elements of said candidate subset **[(col. 9, line 30-44** *"Hyperlinks can be found in homogeneous or heterogeneous corpora. In the domain of academic papers, all papers are semantically similar objects. All citations thus have the same format, from one paper to another. The same is true for patents. In contrast, at least two kinds of hyperlinks are seen on the Web. The first type is navigational hyperlinks, which are intra-site and assist better browsing of the site. The second type reflects endorsement hyperlinks, which are typically across sites. Endorsement links are often associated with pages that are called "resource pages" or "hubs" on the Web, ... These resource pages point to many pages of similar topics and give important information for automatic classification.")]*; thereafter

h) outputting, said second tentative labels as permanent labels associated with said elements of said candidate subset of web pages. **[(col. 9, line 54-55** *"... Web pages of almost every topic point to Netscape.")]*

Chakrabarti et al. does not explicitly teach **f)** training a second classifier using said first tentative labels.

However, *Tresch et al.* teaches f) training a second classifier using said first tentative labels.

[(3.1 Classifier Training Strategies, right column, first paragraph, page 269 *"Incremental training is very efficient when adding the least confident documents ... documents per iteration were incrementally added to the training data."*)] It would have been obvious

at the time the invention was made to a person having ordinary skill in the art to which said

subject matters pertains, to employ an algorithm that first trains a classifier with training data,

and then train a new (second) classifier using the using the results from first training set.

Regarding Claim 2:

Tresch et al. teaches,

h) deciding whether to retrain said second classifier with said second tentative labels. **[(3.1**

Classifier Training Strategies, right column, page 267 *"...Quick (re)training is an ability that is crucial for* any *classifier* ... "*)] It would have been obvious at the time the invention was

made to a person having ordinary skill in the art to which said subject matters pertains, to retrain

the second classifier with the second tentative labels because the second tentative labels reflects

the same labeling.

Regarding Claim 3:

Tresch et al. teaches,

f) training the second classifier using said second tentative labels. **[(3.1 Classifier Training**

Strategies, left column, bottom of page, page 269 *"...Step 1: ... Step 2: ... "*)] It would have

been obvious at the time the invention was made to a person having ordinary skill in the art to

which said subject matters pertains, to retrain the second classifier with the second tentative

labels because the second tentative labels reflects the same labeling.

Regarding Claim 4:

Tresch et al. teaches,

g) collecting said permanent labels associated with said elements of said candidate subset of web

pages [**(3.1 Classifier Training Strategies, left column, first paragraph, page 269** *"These data*

sets have randomly been selected as subsets of a large collection of training documents.")];

h) retraining said first classifier in response to said permanent labels pages. [**(3.1 Classifier**

Training Strategies, left column, second paragraph, page 269 *"One common way is to use*

an incremental training strategy ...The classifier is retrained with the extended set")] It would

have been obvious at the time the invention was made to a person having ordinary skill in the art

to which said subject matters pertains, to employ an algorithm that first trains a classifier with

training data, and then train a new (second) classifier using the using the results from first

training set.

Regarding Claim 5:

Tresch et al. teaches,

second classifier treats selected first regularities differently than said first classifier treats said

first regularities such that said second regularities contradict said first regularities. [**(2.1 Defining**

the Classifier's Schema, page 264 *"A feature is some identifiable part of a document that*

distinguishes between document classes ... "\begin {...}".") e.g., the difference is in the

extension] It would have been obvious at the time the invention was made to a person having

ordinary skill in the art to which said subject matters pertains, to employ an algorithm that first

trains a classifier with training data, and then train a new (second) classifier using the using the

results from first training set.

Regarding Claim 6:

Tresch et al. teaches,

outputting step further includes ignoring training results of said first classifier. **[(2.1 Defining**

the Classifier's Schema, page 264 *"A feature is some identifiable part of a document that*

distinguishes between document classes ... "\begin {...}".") e.g., the difference is in the

extensions due to the different document's retrieve.] It would have been obvious at the time the

invention was made to a person having ordinary skill in the art to which said subject matters

pertains, to employ an algorithm that first trains a classifier with training data, and then train a

new (second) classifier using the using the results from first training set.

Regarding Claim 7:

Tresch et al. teaches,

outputting step further includes combining training results of said first classifier and said second

classifier. **[(3 The Confidence Measure, Figure 3, page 267-268** *"Figure 3 shows the*

distribution of the confidence for a sample classifier. Each dot represents one of the ~2500 test

files ... Both areas have one row for each of the 47 files types.")] It would have been obvious

at the time the invention was made to a person having ordinary skill in the art to which said

subject matters pertains, to employ an algorithm that first trains a classifier with training data,

and then train a new (second) classifier using the using the results from first training set.

Regarding Claim 8:

Chakrabarti et al. teaches,

In a data processing system, a method for learning and combining global regularities and local

regularities for information extraction and classification, said method comprising the steps of

a) initially providing descriptions of said global regularities to a working database, said global

regularities being patterns which may be found over an entire dataset **[(col. 5, line 53-60**

"According to the present invention, a new document stored in the database 112 and containing

citations to and from other documents is classified by the hypertext classifier 110. Initially,

neighboring documents of the new document are identified. Then, for each document and each

class, an initial probability is determined by the hypertext classifier 110, which indicates the

probability that the document fits a particular class.")]; thereafter

b) identifying a candidate subset of the dataset in which local regularities may be found **[(col.**

5, line 60-65 *"Next, iterative relaxation is performed by the hypertext classifier 110 to identify a*

class for each neighboring document using the initial probabilities. A class is selected by the

hypertext classifier 110 into which the new document is to be classified based on the initial

probabilities and identified classes.")]; thereafter

c) tentatively identifying elements having said global regularities in said to candidate subset to

obtain first tentative labels, said first tentative labels being useful for tagging information having

identifiable similarities [(**col. 8, line 14-27** *"Topic identification is one example of extracting structured information from a semi-structured or unstructured source. This is becoming increasingly important as hypertext is becoming a standard datatype supported by most modern databases or database extenders, such as IBM's Universal Database, Oracle's ConText, Verity, etc. Most of these provide support for unsupervised and supervised categorization. The former is clustering; the latter is classification. An automatic classifier is first provided a topic set (i.e., classes), which may be flat or hierarchical, with sample documents for each topic. Using these, the classifier "learns" the topics, in that the classifier is able to classify documents under particular topics. Later, upon receiving new documents, the classifier finds the best matching topics. "*)]; thereafter

d) attaching said first tentative labels onto said identified elements of said candidate subset [(**col. 10, line 35-41** *"In traditional text classification, the object being classified is a self-contained document. A document is a sequence of terms. During training, documents are supplied to the classifier with attached pre-assigned classes. The classifier develops models for the classes in some form or other. During testing, the classifier assigns classes to previously unseen documents."*)]; thereafter

e) employing said attached first tentative labels via one of a class of inductive operations to formulate first local regularities [(**col. 12, line 36-40** *"When a new document is input, the text-based classifier evaluates, using the class models and* Bayes law, *the posteriori probability of the document being generated from each child ... "*)]; thereafter

f) tentatively identifying elements having specific combinations of said global regularities and said local regularities to obtain attached second tentative labels [(**col. 9, line 30-44** *"Hyperlinks*

can be found in homogeneous or heterogeneous corpora. In the domain of academic papers, all

papers are semantically similar objects. All citations thus have the same format, from one paper

to another. The same is true for patents. In contrast, at least two kinds of hyperlinks are seen on

the Web. The first type is navigational hyperlinks, which are intra-site and assist better browsing

of the site. The second type reflects endorsement hyperlinks, which are typically across sites.

Endorsement links are often associated with pages that are called "resource pages" or "hubs" on

the Web, ... These resource pages point to many pages of similar topics and give important

information for automatic classification.")]; thereafter

otherwise

h) employing said second tentative labels via said operation on said candidate subset to

formulate second local regularities, and **i)** repeating from step **f)** until said confidence labels have

been developed.

Chakrabarti et al. does not explicitly teach **g)** testing if estimated error rate is within a pre-

selected tolerance or if a steady state in said attached second tentative labels is evident; and if

true, then rating confidence of said attached second tentative labels and converting selected ones

of said attached second tentative labels to confidence labels upon achieving a preselected

confidence level and then outputting data with said confidence labels;

However, *Tresch et al.* teaches **g)** testing if estimated error rate is within a preselected tolerance

or if a steady state in said attached second tentative labels is evident [**(3.1 Classifier Training**

Strategies, left column, second paragraph, page 267 *"This distribution illustrates the tendency*

of correctly classified files to have a confidence ...to approve the classification of document d as

class c.")]; and if true, then rating confidence of said attached second tentative labels and

converting selected ones of said attached second tentative labels to confidence labels upon

achieving a preselected confidence level and then outputting data with said confidence labels

[(3.1 Classifier Training Strategies, left column, third-fourth paragraph, page 267 *"Figure*

4 illustrates how much feedback can be derived from the confidence measure. ... (see dotted line

hitting threshold).")]; It would have been obvious at the time the invention was made to a person

having ordinary skill in the art to which said subject matters pertains, to employ a system or

method that is designed and constructed to determine the confidence levels using optical

measurements and to generate a output means (graphically) showing and comparing data and

confidence level.

Regarding Claim 9:

Tresch et al. teaches,

said initial global regularity providing step comprises manually inputting descriptions of said

global regularities. **[(3.1 Classifier Training Strategies, left column, second paragraph, page

267** *"To train the classifier, a human expert has to provide a reasonable number of documents*

that are typical of each class ... ")]

Regarding Claim 10:

Chakrabarti et al. teaches,

initial global regularity providing step comprises obtaining said global regularities from a further

one of said class of said inductive operations applied to a subset of said dataset, said subset of

said dataset having been manually labeled. **[(col. 12, line 36-40** "... *using the class model and*

Bayes law, the posterior probability of the document being generated ... **")**

Regarding Claims 11-14:

Due to the repetition of the claim language and logic, claims 11-14 are rejected under the same

rationale as claim 10.

Regarding Claims 15:

Due to the repetition of the claim language and logic, claim 15 is rejected under the same

rationale as claim 1 and claim 8.

Conclusion

10.   The prior art made of record and (listed of form **PTO-892**) not relied upon is considered

pertinent to applicant's disclosure as follows. Applicant or applicant's representative is respect-

fully reminded that in process of patent prosecution i.e., amending of claims in response to a

rejection of claims set forth by the Examiner per Title 35 U.S.C. The patentable novelty must be

clearly shown in view of the state of the art disclosed by the references cited and any objections

made. Moreover, applicant or applicant's representative must clearly show how the amendments

avoid or overcome such references and objections. *See* **37 CFR § 1.111(c).**
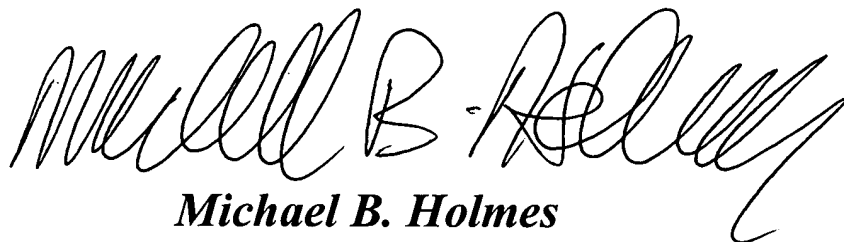
Correspondence Information

11. Any inquiries concerning this communication or earlier communications from the examiner should be directed to **Michael B. Holmes** who may be reached via telephone at **(703) 308-6280**. The examiner can normally be reached Monday through Friday between 8:00 a.m. and 5:00 p.m. eastern standard time.

If you need to send the Examiner, a facsimile transmission regarding After Final issues, please send it to **(703) 746-7238**. If you need to send an Official facsimile transmission, please send it to **(703) 746-7239**. If you would like to send a Non-Official (draft) facsimile transmission the fax is **(703) 746-7240**. If attempts to reach the examiner by telephone are unsuccessful, the **Examiner's Supervisor, Anil Khatri**, may be reached at **(703) 305-0282**.

Any response to this office action should be mailed too:

Director of Patents and Trademarks Washington, D.C. 20231. Hand-delivered responses should be delivered to the Receptionist, located on the fourth floor of **Crystal Park II, 2121 Crystal Drive Arlington, Virginia.**

Michael B. Holmes
Patent Examiner
Artificial Intelligence
Art Unit 2121
United States Department of Commerce
Patent & Trademark Office